Том LXVII | Труды Всесоюзного научно-исследовательского института морского рыбного хозяйства и океанографии (ВНИРО) | 1969

УДК 639.2.001.5

О НЕКОТОРЫХ ТРЕБОВАНИЯХ К СРЕДНИМ В РЫБОХОЗЯЙСТВЕННЫХ ИССЛЕДОВАНИЯХ[1]

С. В. Козлитина

В рыбохозяйственных исследованиях одним из основных при изучении водоемов является метод съемок. По полученным данным строятся карты распределения различных абиотических и биотических показателей, делаются заключения о районах с высокой и низкой продуктивностью, о наличии и концентрации тех или иных организмов и т. д. Как правило, окончательным этапом первичной обработки материалов наблюдений является вычисление средних показателей, характеризующих величину отдельных элементов или определенную их совокупность.

Но средняя величина может быть надежным и объективным показателем лишь тогда, когда при ее вычислении соблюдаются определенные правила. К числу основных из них можно отнести следующие.

Однородность изучаемого материала, которая достигается путем использования данных, относящихся к районам или водоемам с незначительными пространственными контрастами. Так, например, концентрация аммиачного азота в июле 1962 г. на акватории всего Азовского моря изменялась от 14,7 до 85,1 *мг N/м*^{3}, содержание нитритного азота в это же время колебалось от 0 до 48,0 *мг N/м*^{3}.

В первом случае максимальная концентрация превышала минимальную менее чем в 6 раз, во втором — в 48 раз, т. е. можно принять, что полученные данные по аммиачному азоту являются однородными, а по нитритному азоту — неоднородными. И действительно, ошибка средней арифметической для NH_4 оказалась равной 10%, для NO_2 — 55%.

Из этих примеров видно, что вычисление средней концентрации по Азовскому морю возможно только для аммиачного азота. При вычислении среднего содержания нитритного азота необходимо выделить районы с характерными концентрациями.

Исключение систематических ошибок. К ним относятся в основном инструментальные ошибки, влияющие более или менее одинаково на весь ряд наблюдений. Устранять такие ошибки необходимо путем введения поправок. Например, к измеренной температуре воды или воздуха прибавляется поправка из сертификата. При определении численности тюльки в море вводится поправка на уловистость лампары и др.

Нередко устранить инструментальную ошибку невозможно, так как исследователи порой не знают точности своего прибора. К примеру можно привести отбор проб бентоса дночерпателем или зоопланктона сетями Апштейна. Такие наблюдения могут быть использованы для

[1] Работа написана на основании фондовых материалов АзНИИРХа и литературы, приведенной в конце статьи.

изучения изменений, по величине значительно больших, чем предполагаемая ошибка.

Грубые ошибки. При обработке наблюдений приходится считаться с возможностью промахов или внешних влияний, дающих совершенно неверные результаты. Наблюдатель, например, вместо 28 запишет 82, или при измерении скоростей течения что-нибудь попадет под винт, или во время траления трал зацепится за дно. Результаты будут явно искажены и появятся величины, резко отличающиеся от остальных. Казалось бы, что такую величину нужно отбросить, однако подходить к выбраковке подобных данных следует очень осторожно. Все измерения представляют лишь ограниченную выборку из генеральной совокупности, и в выборку может попасть величина, частота которой невелика. И хотя появление ее возможно, принимать ее в расчеты наравне с остальными величинами неправомерно; так как вероятность такой величины близка к нулю, то «выскакивающая» цифра не оказала бы большого влияния на значение математического ожидания, которое вычисляется по формуле

$$M[x] = \sum_{i=1}^{n} x_i p_i ,$$

где p_i — вероятность появления каждого значения.

Следовательно, такие числа можно в исключительных случаях и отбрасывать. Обычно применяют критерии и правила, позволяющие устанавливать принадлежность случайной величины к рассматриваемой совокупности. Наиболее простое и чаще всего встречаемое в практике — правило «трех сигм» (3σ). Если известна величина и среднее квадратичное отклонение σ одного измерения, то находят приближенную ошибку каждого измерения, т. е. отклонения измеренной величины от средней $\varepsilon_i = x_i - \bar{x}$.

Обычно считают маловероятным, чтобы модуль ошибки превышал 3σ. Поэтому если найдется какое-нибудь $|\varepsilon_i| > 3\sigma$, то соответствующее измерение считается содержащим грубую ошибку и отбрасывается. Сигму определяют по формуле

$$\bar{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} x^2{}_i - \overline{x^2} , \qquad (1)$$

где x_i — измеренные величины,

n — число измерений.

После исключения ошибочных данных необходимо среднюю величину и σ вычислить заново. В табл. 1 приведены средние из всех наблюдений (с.в.н) и средние, при вычислении которых исключено лишь по одной недостоверной величине (с.б.н.в.), а также ошибки средних.

Таблица 1

Показатели	Фитопланктон, $г/м^3$	Зоопланктон, $мг/м^3$	Тюлька, шт. на один замет лампары	Нитриты, $мг/м^3$
с.в.н.	1,4	465,0	1656,0	2,3
Ошибка, %	47,0	18,6	21,4	42,1
с.б.н.в.	0,5	347,0	1189,0	1,5
Ошибка, %	12,0	11,7	10,5	34,7

Приведенные данные показывают, что без «выскакивающих» цифры ошибки средних гораздо меньше и средние, следовательно, более точные. Естественно, что и последующие выводы, в основе которых лежат средние, будут также точнее. Например, вычисленная по указанным выше правилам средняя численность тюльки на замет лампары изменилась на 467 шт., что в пересчете на объем моря составит биомассу 467 тыс. *ц*. А это едва ли не наибольший улов тюльки в Азовском море за последние годы.

Случайные ошибки. После исключения систематических и грубых ошибок результаты наблюдений все еще не будут точными. На них действуют различные причины, не поддающиеся учету. Пусть, например, производится взвешивание на точных чувствительных весах. Если в момент отсчета в этой комнате хлопнули дверью, то указатель отклонится в случайном направлении и получится число, отличное от точного. Можно назвать целый ряд случайных причин, приводящих к отклонениям от точного значения. Каждая из этих причин дает малозаметное отклонение, но суммарное воздействие всех причин может дать значительную погрешность. Такого рода ошибки носят название случайных. Устранить их невозможно, и потому, пользуясь закономерностями, характерными для больших совокупностей случайных величин, можно в среднем учесть погрешность наблюдения, вносимую случайными причинами, и степень точности результата наблюдения.

Известно, что случайные ошибки измерения подчиняются нормальному закону распределения, пользуясь которым можно ответить на ряд вопросов, возникающих в измерительной практике, в частности можно говорить о мере точности, средней квадратической вероятной и наибольшей возможной ошибке как отдельного измерения, так и среднего арифметического.

На практике процесс обработки считается законченным, если подсчитана средняя квадратическая ошибка среднего арифметического, которая вычисляется по формуле

$$\sigma_{\overline{x}} = \frac{\overline{\sigma}}{\sqrt{n}},$$

где $\overline{\sigma}$ — средняя квадратическая ошибка одного измерения, определяемая по формуле (1).

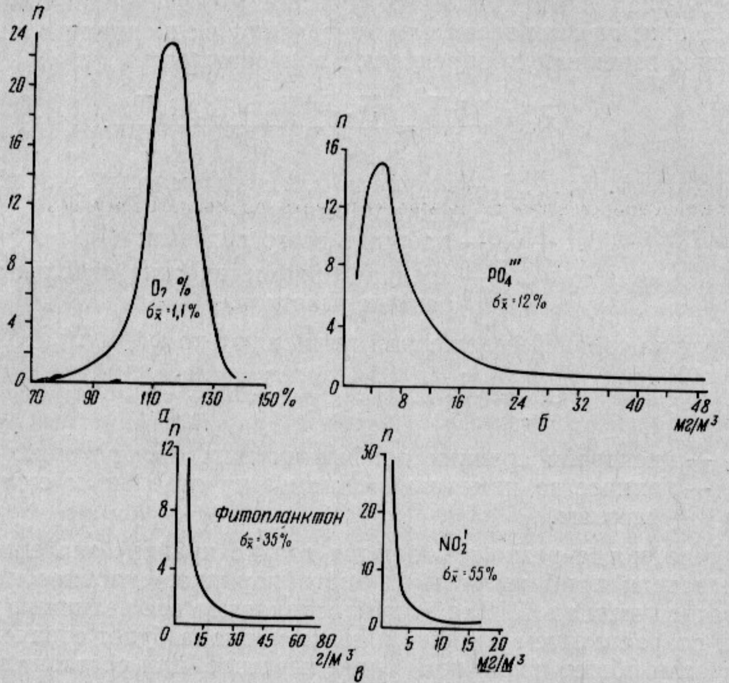
При записи среднего арифметического принято указывать его среднюю квадратическую ошибку, обычно выражаемую в процентах. Замечено, что ошибка находится в тесной связи с характером распределения абсолютных значений изучаемой величины. При симметричном (нормальном) распределении (см. рисунок) ошибка, как правило, наименьшая. При асимметричном γ-распределении ошибка наибольшая (см. рисунок, *в*), достигает 30—50%; при асимметричном распределении, приближающемся к распределению В. И. Смирнова (см. рисунок *б*), ошибка средней в пределах 15—5%.

С целью определения достоверности средних величин, получаемых при рыбохозяйственных исследованиях, проводимых на Азовском море, были подсчитаны средние арифметические, средние квадратические отдельного измерения и среднего арифметического по различным показателям по Таганрогскому заливу и по собственно Азовскому морю, часть из которых приведена в табл. 1 и 2.

Из табл. 2 видно, что ошибки достигают иногда очень больших величин — 30—50%. Нередко это является результатом неверной первичной обработки (не анализировались «выскакивающие» числа), иногда —

большого размаха колебания изучаемой величины. В последнем случае средняя арифметическая не служит характеристикой данного ряда.

Для получения более достоверной средней следует либо увеличить число станций в рейсе, либо пользоваться какой-нибудь другой средней



Кривые распределения абсолютных величин:

а — кислорода, *б* — фосфатов, *в* — нитратов и фитопланктона.

Таблица 2

Ошибка (%) средней концентрации биогенных элементов в поверхностном слое Таганрогского залива и собственно Азовского моря по данным 1962 г.

Месяц	NO_2	NO_3	NH_4	$N_{орг}$	$N_{общ}$	$P_{мин}$	$P_{орг}$	$P_{общ}$	$\delta_i O_2$
Таганрогский залив									
Апрель	21,2	43,9	11,2	9,3	12,9	29,4	13,6	11,8	9,6
Июль	52,4	16,6	17,6	8,6	7,8	9,8	18,7	17,3	6,4
Октябрь	17,4	—	12,1	10,2	—	22,0	15,9	14,2	10,4
Собственно Азовское море									
Апрель	21,9	37,1	7,8	9,5	11,4	15,5	7,0	6,5	7,2
Июль	44,7	15,1	5,4	5,5	6,1	17,7	9,6	13,9	4,1
Октябрь	30,8	10,1	10,7	10,8	—	12,3	5,5	5,7	18,7

(например, медианой, средней взвешенной). Количество станций, необходимое для получения средней, с определенной точностью можно подсчитать по формуле

$$n = \left[\frac{\sigma \cdot 100}{\bar{x}\sigma_{\bar{x}}\%}\right]^2.$$

Однако следует отметить, что незначительное уменьшение ошибки средней влечет резкое увеличение числа станций, поэтому часто рациональнее пользоваться средней взвешенной величиной. Для этого на план изучаемого водоема наносят измеренные величины и проводят изолинии. Затем весь план и отдельные его участки между изолиниями взвешивают на аналитических весах или определяют их площади планиметром.

Среднюю взвешенную определяют по формуле

$$\overline{x}_{b3}=\frac{\Sigma f_i\,\overline{x}_i}{F}=\frac{f_1\overline{x}_1+f_2\overline{x}_2+\ldots+f_n\overline{x}_n}{F},$$

где $f_1, f_2, \ldots, f_n$ — площади или веса (массы) участков между изолиниями;

$F=f_1+f_2+\ldots+f_n$ — площадь всего водоема;

$\overline{x}_i$ — средние арифметические отдельных оконтуренных участков.

Дисперсию средней взвешенной вычисляют по формуле

$$\sigma^2_{\overline{x}_{b3}}=\frac{1}{F^2}\,\Sigma f^2_i\,\sigma^2_{\overline{x}_i}=\frac{1}{F^2}\,\Sigma f^2_i\,\frac{\sigma^2_i}{m_i},$$

где $\sigma^2\overline{x}_i$ — дисперсии средних арифметических оконтуренных участков;

m_i — количество измерений станций по отдельным оконтуренным участкам.

В случае, когда резко отличаются только крайние значения, а средние показатели приблизительно одного порядка, удобнее и надежнее пользоваться медианой. Для этого располагают все величины в убывающем или возрастающем порядке и тогда медиана будет средней по счету цифрой, если общее число измерений нечетное, или составит полусумму двух центральных величин, если количество измерений четное. Например, в Таганрогском заливе в октябре 1962 г. во время съемки были зафиксированы следующие концентрации минерального фосфора (*мг/м*³):

47,6 35,5 21,8 21,4 20,8 18,4 16,9 7,9 7,3 5,5 1,5

Медиана равна 18,4 *мг/м*³.

Из изложенного выше следует, что обработку серии измерений надо проводить в следующем порядке.

1. Определить среднее арифметическое.
2. Найти среднюю квадратическую ошибку отдельного измерения $(\overline{\sigma})$.
3. Определить 3σ и сравнить с ε_i. Если имеются $|\varepsilon_i| \geqslant 3\sigma$, то такие измерения следует отбросить и обработку начать сначала.
4. Определить $\sigma_{\overline{x}}$ среднего арифметического и выразить в процентах.
5. В случае большой $\sigma_{\overline{x}}$ определить среднюю взвешенную.
6. Средняя взвешенная является более надежной характеристикой, так как при определении ее учитывается значимость каждого измерения.

ЛИТЕРАТУРА

Батунер Л. М., Позин М. Е. Математические методы в химической технике. Госхимиздат, 1963.

Великанов М. А. Ошибки измерения и эмпирические зависимости. Гидрометеоиздат, 1962.

Шиголев В. М. Математическая обработка наблюдений. М., Физматгиз, 1962.